Kidney International, Vol. 65 (2004), pp. 2030–2040

A large tandem duplication within the *COL4A5* gene is responsible for the high prevalence of Alport syndrome in French Polynesia

Christelle Arrondel, Georges Deschênes, Yannick Le Meur, Amandine Viau, Christophe Cordonnier, Alain Fournier, Stephane Amadeo, Marie-Claire Gubler, Corinne Antignac, and Laurence Heidet

Inserm U574, Hôpital Necker-Enfants Malades, Université René Descartes, Paris, France; Service de Néphrologie Pédiatrique, Hôpital Armand-Trousseau, Paris, France; Service de Néphrologie, Hôpital universitaire Dupuytren, Limoges, France; Service de Néphrologie, Centre Hospitalier territorial, Papeete, Tahiti, Polynésie Française; Banque d'ADN de Polynésie, Papeete, Tahiti, Polynésie Française; and Service de génétique, Hôpital Necker-Enfants Malades, Paris, France

A large tandem duplication within the *COL4A5* gene is responsible for the high prevalence of Alport syndrome in French Polynesia.

Background. The prevalence of X-linked Alport syndrome, a progressive inherited nephropathy associated with mutations in the type IV collagen gene *COL4A5*, is remarkably high in French Polynesia.

Methods. A vast clinical, genealogic, and molecular study was undertaken in Polynesia, based on public records, patients' interviews, linkage analysis, and mutation screening.

Results and Conclusions. We show that the high frequency of Alport syndrome in this region is due to a founder mutation that occurred onto a common haplotype shared by affected and unaffected individuals, the presence of which precludes indirect molecular diagnosis. We have characterized the mutation as a tandem duplication of 35 *COL4A5* exons, resulting in a ~65% increase in the length of the collagenous domain of the α5(IV) chain, which is still able to assemble into type IV collagen network as shown by immunofluorescence analysis. That mutation is associated with severe and highly penetrant ocular symptoms and with uniformly thin glomerular basememnt membrane (GBM) in male adult patients. However, the rate of progression of the renal disease is very variable from one male patient to another, demonstrating the importance of strong modifier factors. Our results suggest that the 20% to 50% of "missing" *COL4A5* mutations in X-linked Alport syndrome may be rearrangements similar to that reported here, which was not detectable by sequencing of either individual *COL4A5* exons or overlapping cDNA fragments. Finally, we provide the basis for a polymerase chain reaction (PCR) assay that accurately identifies female carriers and allows adequate genetic counseling in this population.

Key words: Alport syndrome, French Polynesia, founder mutation, tandem duplication, molecular diagnosis, genetic counseling.

Received for publication September 18, 2003
and in revised form November 27, 2003
Accepted for publication January 8, 2004

Alport syndrome is an inherited hematuric nephropathy characterized by irregular thinning, thickening and splitting of the glomerular basement membrane (GBM) that leads to end-stage renal disease (ESRD), frequently accompanied by sensorineural deafness and ocular abnormalities (anterior lenticonus and macular flecks) [1–4]. The pathogenesis of the disease has been linked to defect of type IV collagen, which is the major structural component of the basement membranes. Each type IV collagen molecule is composed of three α chains that share a very similar primary structure: (*1*) an approximately 25 residue "7S" domain at the amino terminus; (*2*) a long collagenous domain of approximately 1400 Gly-X-Y repeats, which forms the triple helix together with two other α chains; and (*3*) an approximately 230 residue noncollagenous (NC1) domain at the carboxyl terminus, which is folded into a globular structure. The 7S and NC1 domains are cross-linking domains. Alport syndrom is associated with mutations in any of the *COL4A3*, *COL4A4,* or *COL4A5* genes, encoding the α3, α4, and α5(IV) chains [5, 6]. These three chains assemble into a α3.α4.α5 collagen network, which is extensively cross-linked by disulfide bonds and seems essential for long-term stability of the GBM [7, 8]. About 85% of the affected families exhibit X-linked inheritance of the disease [5]. This most common form of Alport syndrome, estimated to affect 1 in 5000 to 10,000 males worldwide [9], is caused by mutations in the X-chromosomal *COL4A5* gene, at Xq22. To date, more than 300 mutations in *COL4A5* have been identified [4, 10–12]. *COL4A5* is comprised of 53 coding exons [12, 13] spread over ~260 kb of genomic DNA. Major gene rearrangements, detected by Southern blot analysis, are thought to account for 5% to 15% of the mutations in X-linked Alport syndrome [14, 15]. Searches for small mutations, mainly based on

single-strand conformational polymorphism (SSCP) analysis of polymerase chain reaction (PCR) amplified exons followed by nucleotide sequencing of abnormally migrating PCR products, result in a low mutation detection rate, varying from 37% to 50% [10, 16, 17]. These mutations are distributed over the entire length of the gene. Direct DNA sequencing of PCR amplified *COL4A5* exons, using oligonucleotides located further in introns, was reported to enhance the mutation detection rate to 82% [12].

ESRD develops in virtually all affected males with X-linked Alport syndrome. Depending on the rate of progression to ESRD, two types of Alport syndrome have been distinguished [9]: (*1*) a progressive or "juvenile" type in which ESRD occurs around the age of 20 years, always before 31 years, and the course is highly stereotyped within a given family, and (*2*) a "nonprogressive" or "adult" type in which age at ESRD is around 40 years, and the course is much more diverse, making individual prognostication impossible.

The prevalence of Alport syndrome in French Polynesia, which comprises 118 islands and atolls strewn across the eastern South Pacific, is much higher than in any other part of the world. In this paper, we have investigated the clinical epidemiology and molecular genetics of Alport syndrome in French Polynesia.

METHODS

Patients

A large clinical and genealogic study was undertaken in French Polynesia by Y.L. and G.D. in the early 1990s, based on patient and family interviews, medical, church, and state administration records. One thousand and five hundred individuals were investigated. Blood samples were collected from 83 individuals, including 76 with a known Alport syndrome status (11 affected males, 23 unaffected males, 35 affected females and 7 unaffected females). Informed consent was obtained for all tested individuals or their parents. A skin biopsy was performed in an adult male patient, and skin fibroblasts were cultured as previously described [18].

Linkage and haplotype analysis

DNAs were genotyped at four polymorphic microsatellite loci covering a ~6 megabases interval at Xq22. Markers *DXS1191* and *DXS8112* [19] are centromeric to *COL4A5*, *2B6* is located within the *COL4A5* gene [20] and *DXS456* [21] is telomeric to *COL4A5*. Linkage analysis was carried out using the FASTLINK package, assuming an X-linked inheritance, a penetrance of 100% in males and 95% in females [22], a gene frequency of 0.0001, and allowing 0.001 of phenocopies.

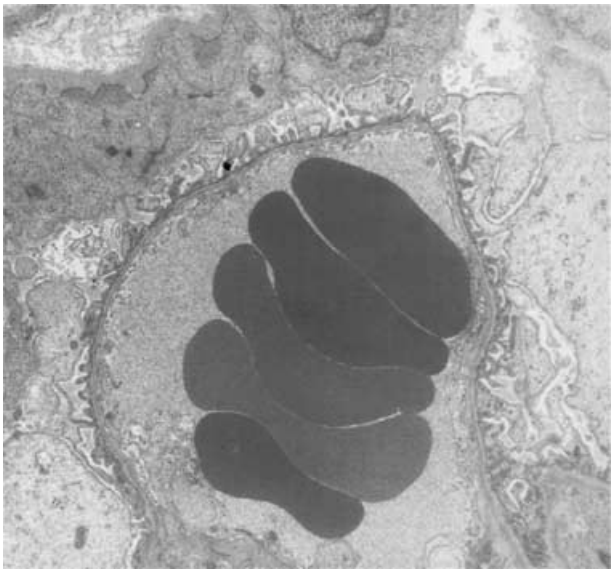


Fig. 1. Electron microscopy from a 33-year-old Tahitian male patient with hematuria, proteinuria and renal failure. A regular thinning of the glomerular basement membrane is observed in one preserved glomerulus (uranyl acetate-lead citrate, ×5000).

DNA sequencing, Southern blot, and pulsed-field gel electrophoresis (PFGE) analysis, COL4A5 probes

All coding exons and the 820 bp of the minimal promoter of *COL4A5* were PCR-amplified using intronic primers as in [12], from the DNAs of two affected males. PCR products were purified by means of the Wizard PCR prep DNA-purification system (Promega, Charbonnière, France) and directly sequenced in both directions by use of an Applied Biosystems DNA sequencer (model 373A; Foster City, CA, USA) and the BigDye terminator cycle sequencing kit (Perkin-Elmer, Courtaboeuf, France), according to the manufacturer's instructions. For Southern blot analysis, DNA was digested and hybridized as previously described [14]. High-molecular-weight DNA was isolated from cultured fibroblasts, and PFGE analysis was carried out as in [23]. *COL4A5* cDNA probes were previously described [24, 25]. Several genomic probes located at regular intervals in *COL4A5* introns 1 and 36 were PCR amplified from three PAC clones covering the *COL4A5* gene (RPCI library clones 149D17, 740A11, and 24A23), which are part of the *Homo sapiens* chromosome X genomic contig (NT_011765) and were purchased from the Sanger Institute.

RNA extraction, reverse transcription (RT)-PCR, and Northern blot analysis

RNA from cultured fibroblasts were extracted using the Rneasy kit (Qiagen, Courtaboeuf, France). cDNA synthesis and PCR amplification were performed as in [23]. Seven overlapping cDNA fragments, covering the entire *COL4A5* cDNA and the poly-A tail, were

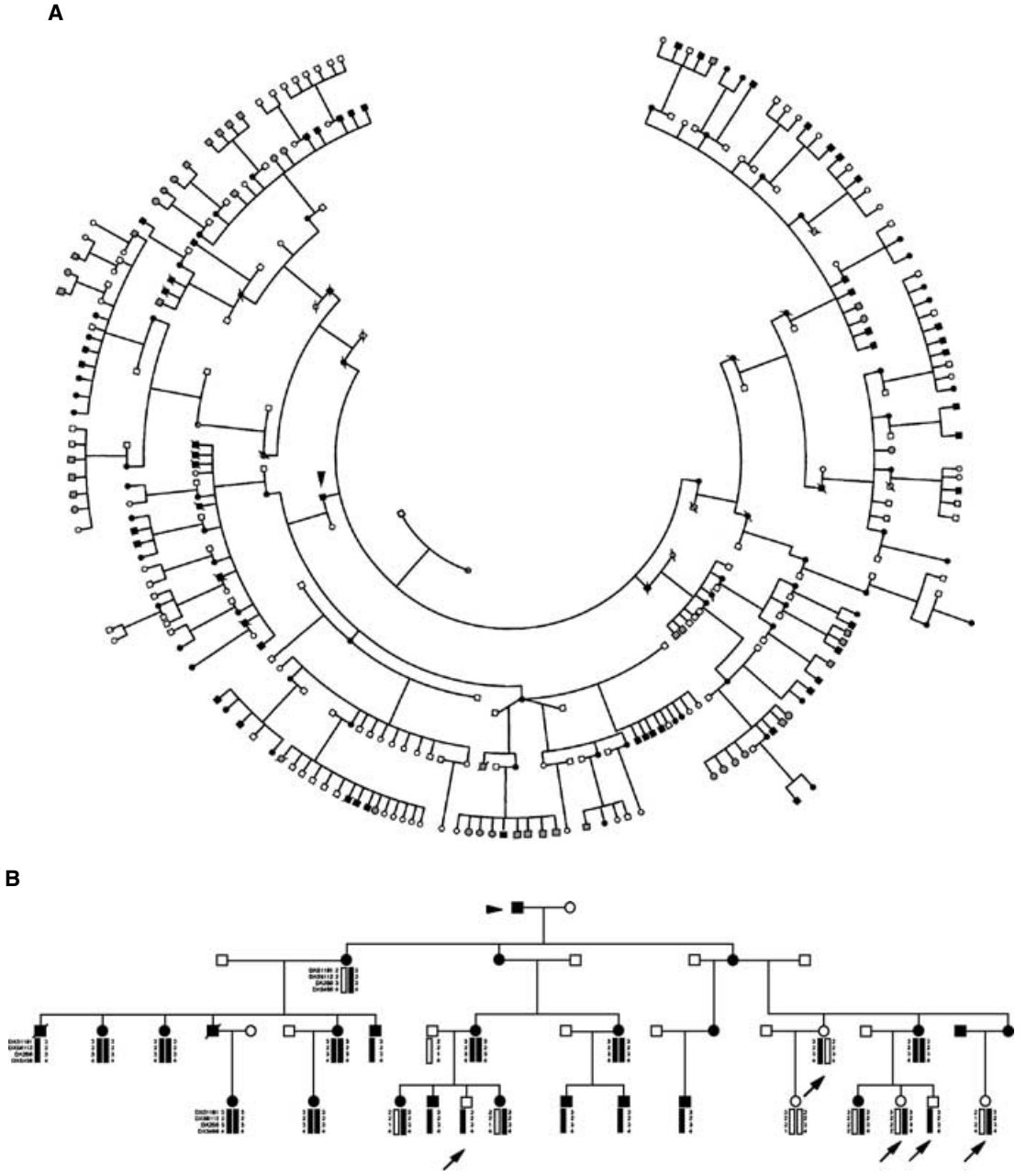


Fig. 2. Pedigree of family I and haplotypes of some genotyped family members from one branch of family I. (*A*) Open squares/circles represent unaffected individuals, black squares/circles represent affected individuals, and gray squares/circles represent individuals with an unknown clinical status. Open circles with a central dot represent obligate carriers with an unknown clinical status. (*B*) The branch of family I (branch 13) descends from the affected son of S-L indicated by the arrowhead. Only genotyped individuals from branch 13 are shown. Haplotypes are indicated for 4 polymorphic microsatellite loci, covering a ~6 megabases interval at Xq22. The marker *2B6* is located within the *COL4A5* gene (20). All affected individuals carry the same haplotype at the *COL4A5* locus. However, some unaffected individuals also carry that haplotype (arrows).

amplified using seven pairs of primers (located in exons 1–13, 12–24, 23–30, 29–37, 36–44, 43–50 and 49–51, respectively) and sequenced. Poly-A RNA were prepared using the mRNA purification kit (Amersham, Brauschweig, UK) and Northern blot analysis was performed using 1 μg of poly-A RNA, and hybridized with the Pc4b *COL4A5* cDNA probe [25].

RESULTS

Clinical data

Clinical investigations showed a remarkably high prevalence of Alport syndrome in French Polynesia. An inventory of 160 Alport syndrome patients (87 females and 73 males) from different islands was made. The

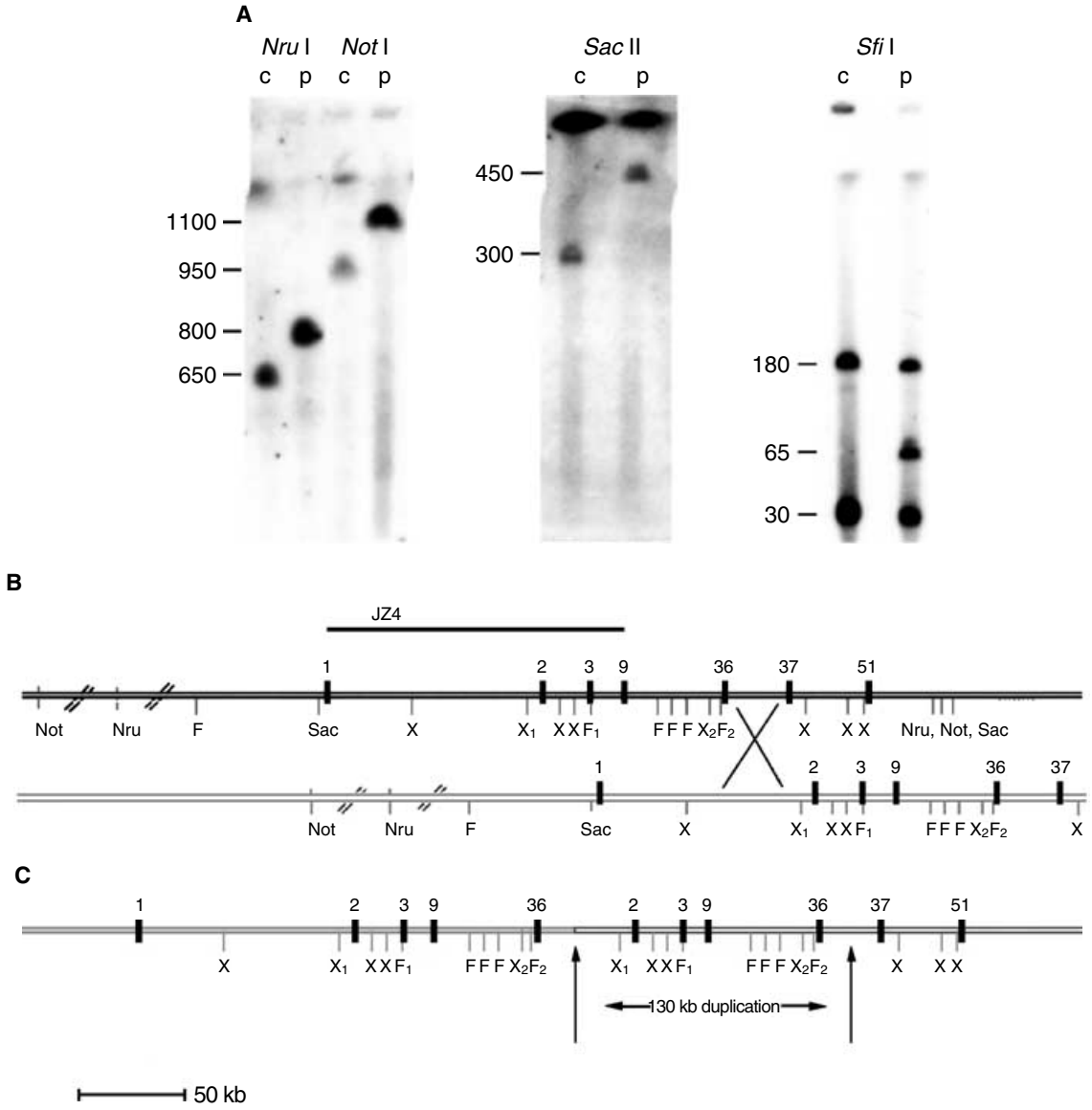
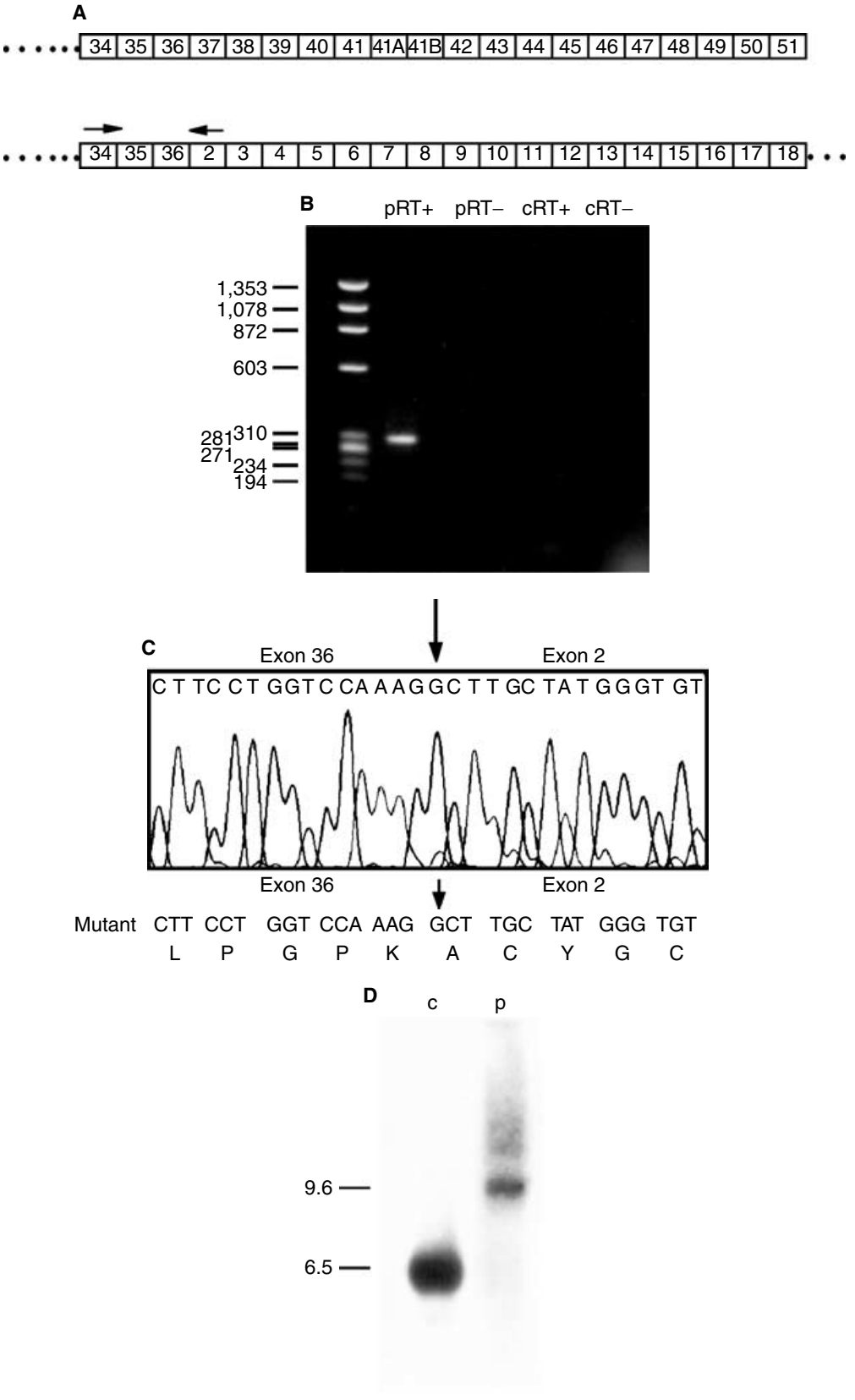


Fig. 3. Analysis of the restriction fragment patterns found in a control individual (c) and a Polynesian patient (p), restriction map of the region and illustration of the proposed mechanisms for the recombination event resulting in the mutated allele. (*A*) After digestion with *Nru*I, *Not*I, or *Sac*II and hybridization with any *COL4A5* cDNA probe, restriction fragments were ~150 kb larger in Polynesian patients than in control DNA. After digestion with *Sfi*I and hybridization with JZ4, a cDNA probe covering exons 1 to 9, a ~65 kb fragment was observed in a Polynesian patient in addition to the normal restriction fragments. (*B*) The long range restriction map around the *COL4A5* gene is from (23); Note: *Not*I, Nru: *Nru*I, Sac: *Sac*II, X: *Xho*I, F: *Sfi*I. *COL4A5* exons 1, 2, 3, 9, 36, 37, and 51 are shown by black boxes. For clarity, the other exons are not indicated. The cross schematizes unequal crossing-over between *COL4A5* introns 1 and 36. (*C*) Schematic representation of the mutant *COL4A5* allele resulting from unequal recombination, containing a tandem gene duplication of ~130 kb, extending from intron 1 to intron 36. We used several probes located at regular intervals in intron 1 and in intron 36 to hybridize *Sfi*I-digested DNA, and thereby mapped the most proximal region of intron 1 hybridizing the abnormal 65 kb fragment to between nt 70,960 and 73,957 of intron 1, and the most distal region of intron 36 hybridizing the abnormal 65 kb fragment to between nt 14,259 and 25,934 of intron 36. This was consistent with the size of an *Sfi*I junction fragment bounded by the *Sfi*I site in intron 35 (5′) (F2), and the *Sfi*I site in exon 3 (3′) (F1). After *Xho*I digestion, a probe covering nt 74,556 to 75,175 of intron 1 and a probe containing exon 36 hybridized, in addition to the normal fragments, an abnormal 40 kb fragment (not shown), in agreement with a junction fragment bounded by *Xho*I sites in intron 31 (5′) (X2) and in intron 1 (3′) (X1).

diagnosis of Alport syndrome was based on hematuria ($N = 151$), association of end-stage renal failure (ESRF), and family history suggesting Alport syndrome ($N = 5$), or solely on evidence of anterior lenticonus ($N = 4$). Alport syndrome was the cause of ESRF in 21% (17 out of 80) of the patients undergoing dialysis or having received a kidney graft in 1991, and in 18% (37 out of 200) in 2003. Age at ESRD in affected males (30 male patients have reached ESRD) varied from 14 years to 44 years. The median renal survival rate in males was 31 years. In females, the renal survival was 100% until the age of 45 years. Seven female patients have reached


A
34 35 36 37 38 39 40 41 41A 41B 42 43 44 45 46 47 48 49 50 51
34 35 36 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
B
pRT+ pRT− cRT+ cRT−
1,353
1,078
872
603
310
281
271
234
194
C
Exon 36
Exon 2
C T TC C T GGT C CA A AG GC T T GC TA T G GGT GT
Exon 36
Exon 2
Mutant CTT CCT GGT CCA AAG GCT TGC TAT GGG TGT
L P G P K A C Y G C
D
c
p
9.6
6.5

ESRF, between the age of 45 and 60 years. Eighty-one patients (41 males and 40 females) were tested for hearing loss, which was detected in 29 (23 males and 6 females) of them. Although all affected males displayed the cardinal symptoms of Alport syndrome, some unusual clinical and morphologic features were consistently observed. The ocular symptoms were remarkably severe, as anterior lenticonus was present in 100% of male patients reaching ESRD and was responsible for severe visual impairment in half, necessitating surgical lens removal. By comparison, lenticonus was found in only 14% of males in a large European cohort of 250 X-linked Alport syndrome families, mostly without significant consequences for visual acuity [4]. Six renal biopsies (performed in three adult males, two 15-year-old boys, and one 12-year-old-boy) were studied by electron microscopy and showed diffuse and uniform thinning of the GBM (Fig. 1). Five biopsies were studied by immunofluorescence with antibodies directed against the α3, the α4, and the α5(IV) collagen chains, respectively, and showed a normal GBM staining for the three chains.

Pedigrees and haplotype analysis

After genealogic investigations, all individuals could be placed in pedigrees of two large families (Fig. 2A, the pedigree of family I) comprising seven generations. The disease was clearly X-linked. It was much more severe in males and there was no father-to-son transmission. Both families originated from Rimatara, a tiny island from the Austral archipelago, 1100 km away from the capital. Family I descends from one individual (S-L), a French whaler who was shipwrecked in 1870 in Tuamotu islands, and settled in Rimatara, where he married a Polynesian lady with whom he had 14 children, four (two females and two males) of whom were affected with Alport syndrome. The study of the distribution of the patients within the different generations of the pedigree showed a consistent increase of the frequency of the disease, with a geometric progression.

DNAs from 83 individuals, including 76 with a known Alport syndrome status, belonging to 4 branches of family I (one branch from each of S-L's affected children) and to family II, were genotyped at four microsatellite loci located on either side of, or within, *COL4A5* at Xq22. Linkage analysis showed a two point maximum LOD score of 7.74 at $\theta_{max} = 0$ between the disease locus and the *COL4A5* intragenic marker *2B6*. Haplotype analysis showed that the same haplotype was carried by all affected males and obligatory carrier females but also by some unaffected individuals (Fig. 2B). This suggested that a *COL4A5* mutation occurred recently onto a haplotype that was already frequent in Rimatara. This finding prevented genetic counseling by haplotype analysis and necessitated the characterization of the mutation in order to provide a diagnosis test.

MUTATION ANALYSIS

Search for small mutation by sequencing and Southern blotting analysis

All 53 coding exons (exons 1 to 51, 41A, and 41B) together with the 820 bp of the minimal promoter of *COL4A5*, amplified from the DNAs of two affected males, were sequenced. This approach did not yield any DNA variants. Southern blot analysis of *Eco*RI- and *Pst*I-digested DNA from an affected male hybridized with overlapping *COL4A5* cDNA probes did not show any abnormal restriction patterns. To look for splicing mutation located outside the immediate exonic regions, RT-PCR was performed using total RNA isolated from skin fibroblast cultures (known to express *COL4A5* [26]) of an affected male. Seven overlapping cDNA fragments, covering the entire *COL4A5* cDNA and the poly-A tail, were amplified and sequenced. No splice variants were detected.

PFGE analysis

Considering the possibility of a complex rearrangement of the gene, we prepared high molecular weight DNA from skin fibroblasts from an affected male and performed PFGE mapping at the *COL4A5* locus. Digestion with *Not*I, *Nru*I, or *Sac*II and hybridization with any *COL4A5* probe showed restriction fragments which were ~150 kb larger in patient than in control fibroblast DNA (Fig. 3A). After digestion of the patient's DNA with *Sfi*I and hybridization with JZ4, covering exons 1 to 9, a ~65 kb fragment was observed in addition to the normal restriction fragments (Fig. 3A). These results suggested that part of *COL4A5* was duplicated, this

Fig. 4. Schematic representation of the 3′ end of the normal (top) and mutated (bottom) *COL4A5* RNAs, reverse transcription-polymerase chain reaction (RT-PCR) amplification of patients (p) and control (c) *COL4A5* fibroblast cDNA with a forward primer located in exon 34 and a reverse primer in exon 2, sequencing analysis of the amplified cDNA in a Polynesian patient, Northern blot analysis of fibroblast RNAs from control (c) and from a male Polynesian patient (p) after hybridization with a *COL4A5* cDNA probe. (*A*) COL4A5 exons are represented by boxes and are numbered. In the mutant RNA, a splicing occurred between exon 36 and exon 2. The locations of the primers used for RT-PCR amplification of the mutant cDNA shown in (B) are indicated by arrows. (*B*) RT-PCR amplification of fibroblast RNA from a male patient generated a 303 bp mutant specific product. RT+ with reverse transcriptase; RT− without reverse transciptase. (*C*) Sequencing of this product showed that nt 3246 in the patient *COL4A5* transcript was followed by nt 82, corresponding to the fusion of exon 36 and exon 2. (*D*) Northern blot analysis showed a transcript of ~6.5 kb in the control and a transcript of ~9.6 kb in the patient.

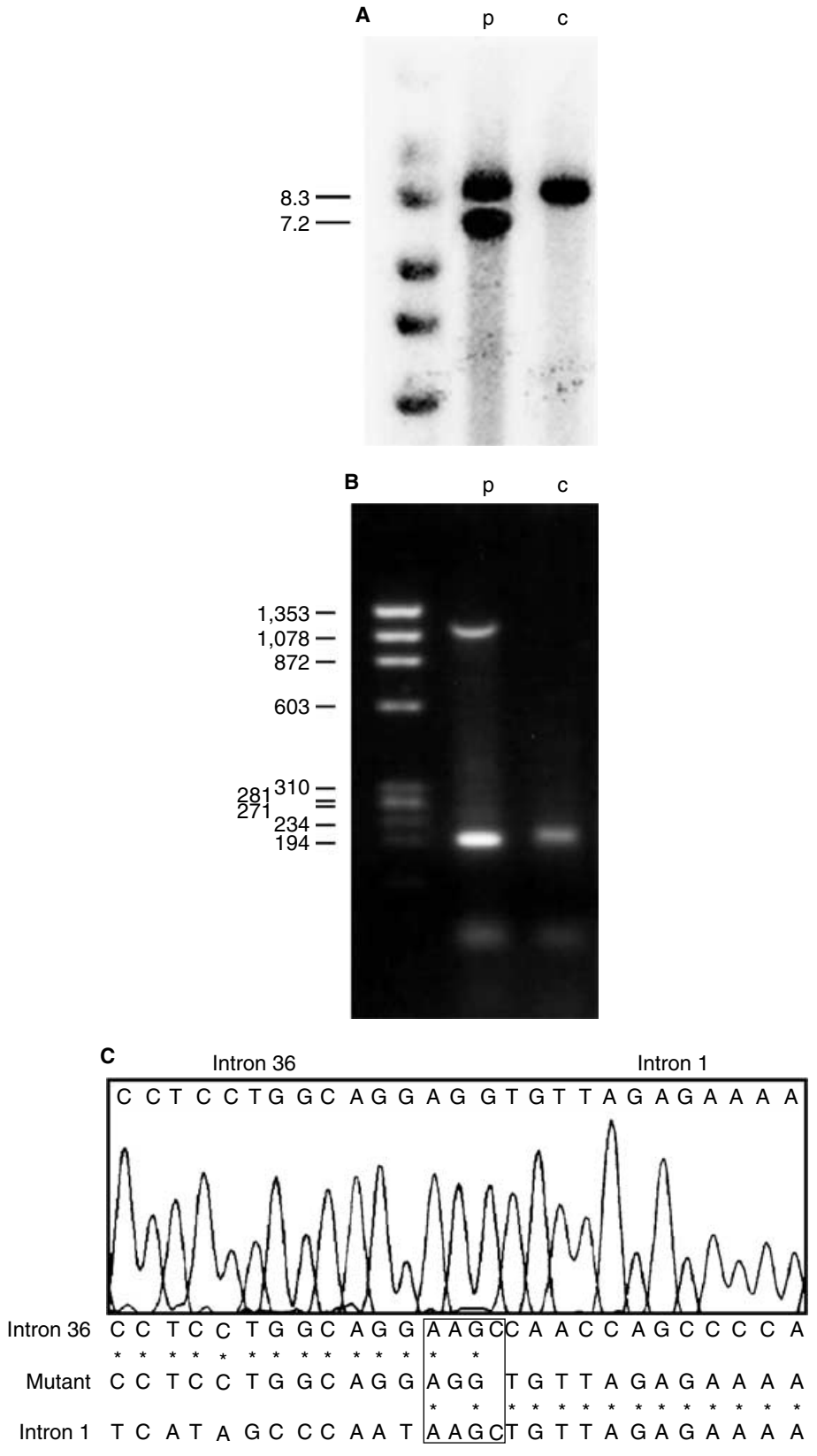


Fig. 5. *EcoR*I restriction pattern of genomic DNA of a patient (p) and control (c), polymerase chain reaction (PCR) amplification of genomic DNA and DNA sequencing analysis of the *COL4A5* mutation in a male Polynesian patient. (*A*) After digestion with *EcoR*I and hybridization with the *COL4A5* int A probe, located at nt 74,556 to 75,175 of intron 1, a 7.2 kb junction fragment was observed in the patient's DNA in addition to the

duplication being localized within the *Sac*II fragment that contains the entire length of *COL4A5* and extends 1.8 kb upstream and 39 kb downstream of the gene (Fig. 3B, top). Since exon 2- (but not exon 1−) and exon 36− (but not exon 37−) probes hybridized the same 65 kb abnormal *Sfi*I fragment in the patient's DNA, and because *COL4A5* introns 1 and 36 are large (99 kb and 29 kb, respectively) and contain several repetitive elements, we hypothesized that a duplication of part of *COL4A5* may have occurred due to unequal crossover between two nonsister chromatids as shown in Figure 3B. We used probes located at regular intervals in intron 1 and intron 36 to hybridize *Sfi*I- and *Xho*I-digested DNA from the male patient. The restriction patterns were in agreement with a ~130 kb duplication including part of intron 1, exons 2–36, and part of intron 36, located in intron 36 (Fig. 3C).

Patient COL4A5 RNA analysis

The exon-intron structure of *COL4A5* shows that exon 36 ends with complete lysine codon and exon 2 starts with complete alanine codon. Thus, in the Tahitian mutated allele as shown in Figure 3C, splicing between exon 36 and exon 2 would maintain the reading frame of the mRNA. To test this hypothesis we performed PCR amplification of a patient's and control fibroblast cDNA using a forward primer in exon 34 and a reverse primer in exon 2 (Fig. 4A, bottom). A 303 bp product was amplified in the patient but not in the control (Fig. 4B). Subcloning and sequencing analysis of this PCR product showed that nt 3246 in the patient's *COL4A5* transcript was followed by nt 82, corresponding to the joining exon 36 and exon 2 (Fig. 4C), with a conservation of the reading frame. RT-PCR using a forward primer in exon 36 and a reverse primer in exon 9 confirmed these data (not shown). Northern blot analysis with poly-A RNA from patient and control fibroblasts hybridized with a *COL4A5* cDNA probe showed a transcript of ~6.5 kb in the control. In the patient, a single transcript of ~9.6 kb was detected (Fig. 4D). This is consistent with a *COL4A5* transcript consisting of 88 exons (exon 1 to 36 followed by exon 2 to 51) (see Fig. 3B, bottom).

Identification of the duplication junction fragment and PCR mutation test

A probe covering nt 74,556 to 75,175 of intron 1 (int A probe) was shown to hybridize a 7.2-kb *Eco*RI abnormal fragment in addition to the expected 8.3 kb fragment, in the patient's DNA (Fig. 5A). All available DNAs were tested, and this fragment was found in all affected males and all obligatory affected females, but not in unaffected individuals carrying the haplotype associated with the Alport syndrome phenotype, showing that the presence of this abnormal fragment was segregating with the disease phenotype. However, two females, a mother and her daughter, considered to be affected because they showed hematuria, did not display the abnormal fragment. These individuals were thought to represent phenocopies. Consequently, clinical reinvestigation was undertaken, which in fact showed that the mother has hematuria due to nephrolithiasis. Further investigations are currently underway to characterize the phenotype in her daughter.

The abnormal 7.2 kb *Eco*RI fragment recognized by the int A probe in intron 1 was shown subsequently to also hybridize to a probe covering nt 14,259 to 14,431 of intron 36. In order to determine precisely the duplication-junction fragment, we performed PCR amplification using a forward primer 5′-CCAGAGAAAATTGAGGAGAG-3′ at nt 14,993 of intron 36 and a reverse primer 5′-ATGTTGAACCATTTAGGAGAA-3′ at nt 71,896 of intron 1. This yielded a 1136 bp product in affected individuals (Fig. 5B). Again, all available DNAs were screened, using a multiplex PCR assay, containing the microsatellite marker *D17S1843* [19] as an internal control. A PCR product was amplified only in individuals previously shown to display the abnormal *Eco*RI fragment with the int A probe. The comparison of this product with the corresponding normal sequences showed that the breakpoint in intron 36 was located at nt 16,073 within a long interspersed nuclear element 1 (L1). The breakpoint in intron 1 was located at nt 71,840, in a 122-bp nonrepeated sequence. Nucleotide alignment around the breakpoint showed a 4 bp junctional homology (AAGC) (Fig. 5C). However, the recombined allele amplified from affected patients showed a (AGG-) junctional sequence instead (Fig. 5C).

DISCUSSION

While Alport syndrome accounts for 1% to 2% of patients reaching ESRD in Europe and 2.3% of the renal transplant patients in the United States [9], 18% of the patients currently undergoing dialysis in French Polynesia, are affected with Alport syndrome. The disease is currently the most prevalent monogenic inherited disorder in this area. In the present paper, we show that the high frequency of Alport in French Polynesia is caused by

expected 8.3 kb fragment. (*B*) Multiplex PCR amplification of genomic DNA with a forward primer located at nt 14,993 of intron 36 and a reverse primer located at nt 71,896 of intron 1 yielded a 1136 bp product in affected individual (p) but not in control (c), whereas the ~200 bp microsatelite marker *D17S1843* was amplified in patient and control. (*C*) Sequencing analysis showed that nt 16,073 of intron 36 is followed by intron 1 sequence found 27,699 nt upstream of exon 2. Sequence of the mutant allele and the corresponding regions of the normal allele in introns 36 and 1 are aligned. The box shows the 4 bp sequence that is identical between intron 36 and intron 1, where the recombination occurred.

a founder mutation that we have characterized here as a tandem duplication of 35 *COL4A5* exons. This rearrangement was not detectable by exon sequencing or conventional Southern blotting using *COL4A5* cDNA probes. Furthermore, due to the size of the cDNA and the fact that the duplication, involving exons 2–36, is located in intron 36, the mutation was not detectable by the sequencing of overlapping *COL4A5* cDNA fragments. In the mutant allele, the duplication results in the transcription of a 9.6 kb RNA comprised of 88 exons (exon 1 to 36 followed by exon 2 to 51), with a conservation of the *COL4A5* reading frame. In the mutated α5(IV) chain, the duplication introduces, within the (Gly-Xaa-Yaa) collagenous domain, a 14 amino acid noncollagenous sequence, including four cysteines from the disulfide rich 7S amino-terminal domain (encoded by the 5′ part of exon 2), followed by 1040 additional amino acids of the collagenous domain (encoded by the 3′ part of exon 2 to exon 36). We [27] and others [28, 29] have shown previously that about one-third of patients affected with X-linked Alport syndrome normally express the GBM α3.α4.α5(IV) collagen network, whereas *COL4A5* mutations are ultimately responsible for a defective assembly of that network in two thirds of the patients. A normal immunohistologic distribution of the collagen chains was reported to be mostly associated with missense or in-frame mutations, whereas large rearrangements were shown to be associated with a lack of α3.α4.α5(IV) network expression [4, 28, 29]. Here, we report a large in frame duplication of *COL4A5* resulting in a 1054 amino acid longer α5(IV) chain (an increase of ~65% in the length of the triple helical domain of the chain) which, surprisingly, is still able to assemble with α3(IV) and α4(IV) into the GBM collagen network, as shown by the normal expression of the α5(IV) chain in the Tahitian patients' kidneys. A similarly mutated α2(I) chain with an aberrantly long, albeit shorter than that we report here, triple helical domain was shown to be incorporated into fibrillar procollagen molecules [30]. Additionally, the fact that the triple helical domains of type IV collagen molecules naturally comprise frequent noncollagenous interruptions, thought to play a role in flexibility [31], may explain why the insertion of a short noncollagenous sequence does not affect the assembly of the mutated chain.

Isolated populations where affected individuals are recently descended from a mutation-bearing founder usually provide opportunity of studying disease with homogeneous phenotype, due to little variations created by allelic heterogeneity, geographic or ethnic variation. Surprisingly, the progression of the renal disease in Polynesian patients, who all bear the same *COL4A5* mutation, varies a lot from one male to another, making it impossible to classify the disease as either the juvenile or the adult form of Alport syndrome. This suggests that strong environmental and/or genetic factors modify the progression of the Alport syndrome nephropathy. It will be interesting, in a first attempt, to check whether the level of transcription of the mutant *COL4A5* gene in skin fibroblasts correlates with a slower progression of the disease.

Anterior lenticonus in X-linked Alport syndrome occurs more frequently in patients carrying large *COL4A5* deletions, nonsense or frameshift mutations, usually associated with a lack of expression of α5(IV)-containing type IV collagen networks [4]. However, ocular symptoms were present in all Tahitian males reaching ESRD, suggesting that the mutant protein is directly responsible for severe abnormalities of the anterior lens capsule, which normally express all of the α(IV) chains [32]. Another striking feature was the diffuse and uniform thinning of the GBM in adult Tahitian patients. Diffuse thinning of the GBM is observed in patients affected with benign familial hematuria, which is associated with heterozygous *COL4A3* or *COL4A4* mutations [33, 34]. It may be the only morphologic defect observed in children affected with Alport syndrome, but, when present in adult Alport syndrome patients, thinning of the GBM is most often segmental and associated with thickening and splitting. Our data further demonstrate that "thin basement membrane" is not necessarily associated with good renal prognosis.

The mechanism responsible for the tandem duplication we report here is likely to be a non-homologous recombination due to interchromosomal misalignment. Recombination between homologous DNA sequences during meiosis after chromosomal pairing, breakage and crossover is an essential cellular process. However, nonequal chromosomal pairing resulting in nonhomologous recombination is a source of harmful deletions and duplications and diseases [35]. Hot spots of homologous recombination between nonallelic, interspersed, or tandem repeats, or between members of gene families or pseudogenes, have been observed at the duplication and deletion breakpoints in a number of human diseases [35]. In several reported cases, however, break points occur through nonhomologous end-joining between repeat-free DNA fragments, or between a repeat and a repeat-free sequence [36], as is the case in the duplication we report here. The involvement of short direct repeats of 2 to 8 bp at the duplication/deletion break points, such as the 4 bp (AAGC) junctional homology which was found here, has been well documented [37–40]. However, the recombined allele amplified from Tahitian patients showed a (AGG-) junctional sequence instead of the (AAGC) sequence. This is consistent with minor deletions and/or insertions often found at the site of rearrangements that possibly originate from repair processes during which free ends of DNA strands are filled in prior to ligation [41]. Although less common than recombination due to exonic sequence homology [42–45], nonhomologous recombination involving introns has

been previously reported in fibrillar [30] and nonfibrillar [46] collagens. While we were not able to precisely date the occurrence of this mutation, it must have occurred before 1870, and the two reported families are probably related through the wife of S-L, the French whaler. Several factors may have led to the progression of the prevalence of the disease in French Polynesia: the delayed onset of renal failure in some males, thereby allowing them to procreate, and the local demographics with a fertility rate of 5.3 children born per woman at the time the clinical study was undertaken. In most patients affected with X-linked Alport syndrome, the absence (in males) or focal distribution (in females) of the α5(IV) chain, in the GBM as well as in the dermal basement membrane can be used for rapid and definitive diagnosis, by immunofluorescence analysis of a simple skin biopsy [27]. However, in Polynesian patients, the synthesis and incorporation of the mutated α5(IV) chain into the type IV collagen network prevents such diagnosis. In addition, we have shown that the Xq22 haplotype associated with the disease phenotype is also present in unaffected individuals, precluding any indirect molecular diagnosis. Therefore, the mutation analysis and the multiplex PCR-based test we report here provides the basis for a rapid, simple, and robust test for Alport syndrome in this area. It will be especially useful for detection of female carriers and adequate genetic counseling.

The *COL4A5* mutation rate reported in different large cohort of X-linked Alport syndrome patients [10, 16–17] is low, suggesting that many mutations are overlooked. Our report suggests that the 20% to 50% of "missing" mutations in X-linked Alport syndrome may be large *COL4A5* rearrangements, which cannot be detected by sequencing of genomic DNA and may escape sequencing analysis of overlapping cDNA fragments. Families for which no mutations are found by sequencing analysis should therefore be tested by PFGE restriction mapping.

ACKNOWLEDGMENTS

We are grateful to the patients and their families for their participation in this work. We thank Professor A. Bourdais for initiating the clinical study. We thank Mélanie Jouin for excellent technical assistance. The help of Olivier Gribouval for LOD score analysis is gratefully acknowledged. We are very grateful to Dr. Eileen Boye and Dr. Ernie Esquivel for their thoughtful criticisms of the manuscript. This work was supported by the Association Française contre les Myopathies, the Association pour l'Utilisation du Rein Artificiel, and the Association Claude Bernard.

Reprint requests to Laurence Heidet, Inserm U574, Tour Lavoisier 6ème étage, Hôpital Necker-Enfants Malades 75743 Paris Cedex 15, France.
E-mail: heidet@necker.fr

REFERENCES

1. Habib R, Gubler MC, Hinglais N, *et al*: Alport's syndrome: Experience at Hopital Necker, *Kidney Int* 11:S20–S28, 1982
2. Grunfeld JP: The clinical spectrum of hereditary nephritis. *Kidney Int* 1:83–92, 1985
3. Kashtan CE, Michael AF: Alport syndrome. *Kidney Int* 50:1445–1463, 1996
4. Jais JP, Knebelmann B, Giatras I, *et al*: X-linked Alport syndrome: Natural history in 195 families and genotype- phenotype correlations in males. *J Am Soc Nephrol* 11:649–657, 2000
5. Barker DF, Hostikka SL, Zhou J, *et al*: Identification of mutations in the *COL4A5* collagen gene in Alport syndrome. *Science* 248:1224–1227, 1990
6. Mochizuki T, Lemmink HH, Mariyama M, *et al*: Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome. *Nat Genet* 8:77–81, 1994
7. Gunwar S, Ballester F, Noelken ME, *et al*: Glomerular basement membrane. Identification of a novel disulfide-cross-linked network of alpha3, alpha4, and alpha5 chains of type IV collagen and its implications for the pathogenesis of Alport syndrome. *J Biol Chem* 273:8767–8775, 1998
8. Kalluri R, Shield CF, Todd P, *et al*: Isoform switching of type IV collagen is developmentally arrested in X-linked Alport syndrome leading to increased susceptibility of renal basement membranes to endoproteolysis. *J Clin Invest* 99:2470–2478, 1997
9. Atkin CL, Gregory MC, Border WA: Alport syndrome, in *Diseases of the Kidney*, edited by Schrier RW, Gottschalk CW, Boston, Little, Brown, 1988, pp 617–641
10. Renieri A, Bruttini M, Galli L, *et al*: X-linked Alport syndrome: An SSCP-based mutation survey over all 51 exons of the COL4A5 gene. *Am J Hum Genet* 58:1192–1204, 1996
11. Lemmink HH, Schroder CH, Monnens LA, Smeets HJ: The clinical spectrum of type IV collagen mutations. *Hum Mutat* 9:477–499, 1997
12. Martin P, Heiskari N, Zhou J, *et al*: High mutation detection rate in the *COL4A5* collagen gene in suspected Alport syndrome using PCR and direct DNA sequencing. *J Am Soc Nephrol* 9:2291–2301, 1998
13. Zhou J, Leinonen A, Tryggvason K: Structure of the human type IV collagen *COL4A5* gene. *J Biol Chem* 269:6608–6614, 1994
14. Antignac C, Knebelmann B, Drouot L, *et al*: Deletions in the COL4A5 collagen gene in X-linked Alport syndrome. Characterization of the pathological transcripts in nonrenal cells and correlation with disease expression. *J Clin Invest* 93:1195–1207, 1994
15. Tryggvason K: Mutations in type IV collagen genes and Alport phenotypes. *Contrib Nephrol* 117:154–171, 1996
16. Kawai S, Nomura S, Harano T, *et al*: The *COL4A5* gene in Japanese Alport syndrome patients: Spectrum of mutations of all exons. The Japanese Alport Network. *Kidney Int* 49:814–822, 1996
17. Knebelmann B, Breillat C, Forestier L, *et al*: Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport syndrome. *Am J Hum Genet* 59:1221–1232, 1996
18. Bourgeron T, Chretien D, Amati P, *et al*: Expression of respiratory chain deficiencies in human cultured cells. *Neuromuscul Disord* 3: 605–608, 1993
19. Dib C, Faure S, Fizames C, *et al*: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–154, 1996
20. Barker DF, Cleverly J, Fain PR: Two CA-dinucleotide polymorphisms at the COL4A5 (Alport syndrome) gene in Xq22. *Nucleic Acids Res* 20:929, 1992
21. Luty JA, Guo Z, Willard HF, *et al*: Five polymorphic microsatellite VNTRs on the human X chromosome. *Am J Hum Genet* 46:776–783, 1990
22. Jais JP, Knebelmann B, Giatras I, *et al*: X-linked Alport syndrome: Natural history and genotype-phenotype correlations in females belonging to 195 families. A "European Community Alport Syndrome Concerted Action" study. *J Am Soc Nephrol* 14:2603–2610, 2003
23. Heidet L, Dahan K, Zhou J, *et al*: Deletions of both alpha 5(IV) and alpha 6(IV) collagen genes in Alport syndrome and in Alport syndrome associated with smooth muscle tumours. *Hum Mol Genet* 4:99–108, 1995
24. Zhou J, Hertz JM, Leinonen A, Tryggvason K: Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. *J Biol Chem* 267:12475–12481, 1992

25. Hostikka SL, Eddy RL, Byers MG, *et al*: Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome. *Proc Natl Acad Sci USA* 87:1606–1610, 1990
26. Sasaki S, Zhou B, Fan WW, Kim Y, *et al*: Expression of mRNA for type IV collagen alpha1, alpha5 and alpha6 chains by cultured dermal fibroblasts from patients with X-linked Alport syndrome. *Matrix Biol* 17:279–291, 1998
27. Gubler MC, Antignac C, Deschênes G, *et al*: Genetic, clinical, and morphologic heterogeneity in Alport's syndrome. *Adv Nephrol* 22: 15–35, 1993
28. Naito I, Kawai S, Nomura S, *et al*: Relationship between COL4A5 gene mutation and distribution of type IV collagen in male X-linked Alport syndrome. Japanese Alport Network. *Kidney Int* 50:304–311, 1996
29. Mazzucco G, Barsotti P, Muda AO, *et al*: Ultrastructural and immunohistochemical findings in Alport's syndrome: A study of 108 patients from 97 Italian families with particular emphasis on COL4A5 gene mutation correlations. *J Am Soc Nephrol* 9:1023–1031, 1998
30. Raff ML, Craigen WJ, Smith LT, *et al*: Partial *COL1A2* gene duplication produces features of osteogenesis imperfecta and Ehlers-Danlos syndrome type VII. *Hum Genet* 106:19–28, 2000
31. Tryggvason K, Soininen R, Hostikka SL, *et al*: Structure of the human type IV collagen genes. *Ann N Y Acad Sci* 580:97–111, 1990
32. Ohkubo S, Takeda H, Higashide T, *et al*: Immunohistochemical and molecular genetic evidence for type IV collagen alpha5 chain abnormality in the anterior lenticonus associated with Alport syndrome. *Arch Ophthalmol* 121:846–850, 2003
33. Boye E, Mollet G, Forestier L, *et al*: Determination of the genomic structure of the *COL4A4* gene and of novel mutations causing autosomal recessive Alport syndrome. *Am J Hum Genet* 63:1329–1340, 1998
34. Heidet L, Arrondel C, Forestier L, *et al*: Structure of the human type IV collagen gene *COL4A3* and mutations in autosomal Alport syndrome. *J Am Soc Nephrol* 12:97–106, 2001
35. Lupski JR: Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* 14:417–422, 1998
36. Stankiewicz P, Shaw CJ, Dapper JD, *et al*: Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *Am J Hum Genet* 72:1101–1116, 2003
37. Henthorn PS, Smithies O, Mager DL: Molecular analysis of deletions in the human beta-globin gene cluster: Deletion junctions and locations of breakpoints. *Genomics* 6:226–237, 1990
38. Hori T, Tomatsu S, Nakashima Y, *et al*: Mucopolysaccharidosis type IVA: Common double deletion in the N-acetylgalactosamine-6-sulfatase gene (GALNS). *Genomics* 26:535–542, 1995
39. Jalanko A, Manninen T, Peltonen L: Deletion of the C-terminal end of aspartylglucosaminidase resulting in a lysosomal accumulation disease: Evidence for a unique genomic rearrangement. *Hum Mol Genet* 4:435–441, 1995
40. Dabora SL, Nieto AA, Franz D, *et al*: Characterisation of six large deletions in TSC2 identified using long range PCR suggests diverse mechanisms including Alu mediated recombination. *J Med Genet* 37:877–883, 2000
41. Thode S, Schafer A, Pfeiffer P, Vielmetter W: A novel pathway of DNA end-to-end joining. *Cell* 60:921–928, 1990
42. Cohn DH, Zhang X, Byers PH: Homology-mediated recombination between type I collagen gene exons results in an internal tandem duplication and lethal osteogenesis imperfecta. *Hum Mutat* 2:21–27, 1993
43. Tiller GE, Rimoin DL, Murray LW, Cohn DH: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual with spondyloepiphyseal dysplasia. *Proc Natl Acad Sci USA* 87:3889–3893, 1990
44. Cabral WA, Mertts MV, Makareeva E, *et al*: Type I collagen triplet duplication mutation in lethal osteogenesis imperfecta shifts register of alpha chains throughout the helix and disrupts incorporation of mutant helices into fibrils and extracellular matrix. *J Biol Chem* 278:10006–10012, 2003
45. Hamalainen ER, Renieri A, Pecoraro C, *et al*: Unequal homologous crossing over resulting in duplication of 36 base pairs within exon 47 of the COL4A5 gene in a family with Alport syndrome. *Hum Mutat* 8:265–269, 1996
46. Vetrie D, Boye E, Flinter F, *et al*: DNA rearrangements in the alpha 5(IV) collagen gene (*COL4A5*) of individuals with Alport syndrome: Further refinement using pulsed-field gel electrophoresis. *Genomics* 14:624–633, 1992